Tree-Ring Watermarks: Fingerprints for Diffusion Images that are Invisible and Robust

Year: 2023 | Citations: 153 | Authors: Yuxin Wen, John Kirchenbauer, Jonas Geiping, T. Goldstein

---

Abstract

Watermarking the outputs of generative models is a crucial technique for tracing copyright and preventing potential harm from AI-generated content. In this paper, we introduce a novel technique called Tree-Ring Watermarking that robustly fingerprints diffusion model outputs. Unlike existing methods that perform post-hoc modifications to images after sampling, Tree-Ring Watermarking subtly influences the entire sampling process, resulting in a model fingerprint that is invisible to humans. The watermark embeds a pattern into the initial noise vector used for sampling. These patterns are structured in Fourier space so that they are invariant to convolutions, crops, dilations, flips, and rotations. After image generation, the watermark signal is detected by inverting the diffusion process to retrieve the noise vector, which is then checked for the embedded signal. We demonstrate that this technique can be easily applied to arbitrary diffusion models, including text-conditioned Stable Diffusion, as a plug-in with negligible loss in FID. Our watermark is semantically hidden in the image space and is far more robust than watermarking alternatives that are currently deployed. Code is available at https://github.com/YuxinWenRick/tree-ring-watermark.